METHODOLOGY FOR ANALYSIS AND CHARACTERIZATION OF NONDESTRUCTIVE INSPECTION CAPABILITY DATA

W. D. Rummel, B. K. Christner and D. L. Long

Martin Marietta Astronautics Group
P.O. Box 179
Denver, Colorado 80201

INTRODUCTION

The need to quantify and validate nondestructive inspection (NDI) performance capabilities has become increasingly necessary with the application of fatigue and fracture mechanics as the basis of design for modern engineering systems. The demand for quantitative characterization of NDI performance capabilities has led to the development of several analytical tools for this purpose.

A method of presenting inspection data as the probability of detection (POD) as a function of flaw size was developed initially as a method of consolidating and presenting inspection capability data in a quantitative manner that was easily understood. Several methods of analysis have now evolved to generate POD curves using data collected experimentally by passing test articles containing cracks of varying sizes through an inspection process and recording the outcome. The method of analysis most appropriate for use depends on the type and quantity of inspection capability data available for analysis. The POD curve provides a convenient method of presenting the effects of varying process parameters on crack detection performance and for the overall detection capabilities of an inspection process.

The traditional POD curve alone however, does not fully describe the capability of an inspection process. The POD curve provides no information on the false call rate of the inspection process being depicted nor the effect of changing the signal amplitude level at which the reject/accept inspection decision is made. The accept/reject decision level has a direct effect on the resulting detection capability and the probability for false alarms. This relationship must be understood before an inspection process can be effectively implemented for a production application at a high confidence level.

INSPECTION RELIABILITY AND CONDITIONAL PROBABILITY

Knowing quantitatively the overall detection capability of an inspection process is of primary importance for design purposes and for establishing production inspection acceptance criteria. The outcome of an inspection process is not a simple accept / reject decision as is commonly thought. The results from an inspection are better described as a case of conditional probability and the inspection itself as a statistical hypothesis test. Instead of a simple accept / reject response, an inspection can produce four possible outcomes as illustrated in Fig. 1 and is a test of the hypothesis that a part rejected by an inspection process does in fact contain a rejectable defect. The desirable results from an inspection process are a true positive response and a true negative response. However the inspection may also produce a false positive/Type II error (false alarm) or a false negative/Type I error (miss) response. Production inspection processes

		STIMULI: POSITIVE	STIMULI: NEGATIVE
RESPONSE	TRUE	TRUE POSITIVE (NO ERROR)	FALSE POSITIVE (TYPE II ERROR)
RESPONSE	FALSE	FALSE NEGATIVE (TYPE I ERROR)	TRUE NEGATIVE (NO ERROR)

Fig. 1. Possible outcomes from an inspection process.

should be validated and operated at that point which maximizes the probability for true positive and true negative responses and limits the probability for Type I and Type II error responses based on the accept/reject criteria for the part being inspected. The probability of a true positive response is not constant for an inspection process however, but varies with the the size, type, orientation, location etc. of the defect being interogated and necessitates the use of additional analysis techniques to fully quantify the performance characteristics of an inspection process.

ANALYSIS AND QUANTIFICATION OF NDI PERFORMANCE DATA

Signal / Noise Analysis

The probability of detecting a crack of a certain size is dependent on the signal (plus noise) and noise distributions generated by application of the inspection technique to flaws of that specific size and the acceptance criteria applied in the decision process. The overall signal distribution is influenced by both flaw to flaw variations between flaws of equal size and variations in signal response resulting from repetitive calibration and application of the inspection process itself.

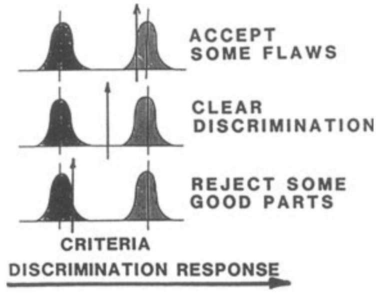

Fig. 2. Relationship of signal and noise distributions with acceptance criteria to determine flaw detection performance.

Inspection Acceptance Decision Criteria

Given the signal and noise distributions for a flaw of a specific size, the inspection accept/reject decision criteria used determines the resultant discrimination capability of the inspection process. Consider an inspection process with a measurable separation in noise and flaw signal responses as shown in Fig. 2. If the acceptance criteria for this inspection (indicated by the vertical arrow) is set too high, some flaws will be accepted (missed). If the acceptance criteria is set at a level that provides clear separation of the noise signal from the flaw signal, a high percentage of flaws will be rejected and few false calls will occur. If the acceptance criteria is set too low, all flaws will be rejected, however some good parts may also be rejected.

The analysis of the relationship between the inspection signal distribution, the noise distribution and the acceptance criteria provides a measure of the detection capability and false call level for flaws of a specific size. This relationship for the range of flaw sizes being considered is used to construct the probability of detection (POD) curve as depicted in Fig. 3.

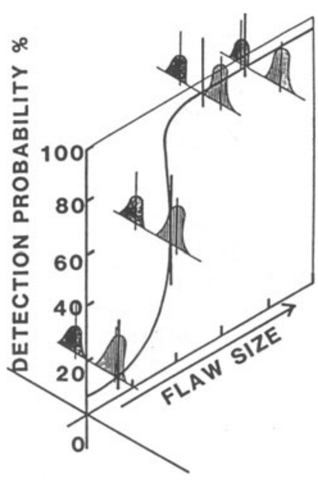

Fig. 3. Interaction of acceptance criteria with the signal and noise distribution to determine probability of detection.

PROBABILITY OF DETECTION (POD) CURVES

The statistic that is of primary interest for production inspections is the probability of detecting flaws exceeding the design critical size. As previously described, performance in terms of positive detection as a function of flaw size is conveniently described by the POD curve. For an inspection process that produces only hit or miss type outputs (ie. liquid penetrant inspection), POD curves are most commonly plotted using the maximum likelihood and the binomial grouping methods of analyses.

Binomial Grouping Analysis

The binomial grouping or moving average method of analysis described by Rummel et al [1] is accomplished by passing a large number of flaws of varying sizes, bracketing the anticipated detection capability of the inspection process, through the inspection and recording the output in terms of detection or failure to detect. The probability of detection as a function of flaw size is obtained by (1) ordering the data from largest flaw size to smallest flaw size; (2) selecting a sample size for binomial statistics that is consistent with the desired reliability and confidence level (for example a sample size of 29 provides a 90% reliablity / 95% confidence level analysis); (3) counting down sequentially starting with the largest flaw to obtain a sample of the selected size; (4) calculating the point estimate for detection probability (ie. the number of detections divided by the sample size) and plotting at the median flaw size in the sample; (5) calculating and plotting the lower confidence limit based on the detection success and the sample size; (6) dropping the largest flaw from the sample and adding the next largest flaw not yet sampled; and (7) repeating the analytical and plotting process until the data are exhausted. Regression analysis techniques may then be used to fit a curve to the resulting point data. The binomial grouping method is appropriate for applications were a large quanitity of data is available covering a relatively wide crack size range.

Maximum Likelihood Analysis

The maximum likelihood method of POD analysis that is most commonly used is based on the logodds or log logistic model for the POD function. This function can be expressed as:

$$\mathrm{POD}(a) = \frac{\exp[\alpha + \beta \ln(a)]}{1 + \exp[\alpha + \beta \ln(a)]} \qquad (1)$$

Where a equals the flaw size.

The parameters α and β for the logodds POD model can be estimated using the principles of maximum likelihood as described by Berens and Hovey [2] and Berkson [3]. Maximum likelihood estimation of the model parameters does not require grouping of the data but is based directly on the observed outcomes of 0 for non-detection and 1 for detection. Given a set of hit-miss data the maximum likelihood method converges to those model parameters which maximize the probability for obtaining the observed data by iterative solution of simultaneous equations (2) and (3).

$$0 = \sum_{i=1}^{n} n_i p_i - \sum_{i=1}^{n} \frac{n_i \exp(\alpha+\beta \ln(a_i))}{1+\exp(\alpha+\beta \ln(a_i))} \quad (2)$$

$$0 = \sum_{i=1}^{n} n_i p_i \ln(a_i) - \sum_{i=1}^{n} \frac{n_i \ln(a_i) \exp(\alpha+\beta \ln(a_i))}{1+ \exp(\alpha+\beta \ln(a_i))} \quad (3)$$

where p_i=1 if the flaw was detected and p_i=0 if the flaw was not detected for single inspection data and p_i equals the proportion of times an individual crack was detected for multiple inspection data; n_i equals the number of times the i^{th} crack was inspected; and a_i equals crack length. A lower confidence bound on the POD function can then be calculated and plotted. The maximum likelihood method permits the use of fewer data points than does the binomial grouping method. However, in cases where the inspection data deviates from the log logistic model the method will not produce a definitive solution.

â Versus a Analysis

For inspection processes that produce quantitative and discrete signal outputs, the POD function can be generated using the relationship between flaw signal (â) and actual flaw size (a). The â versus a method described by Berens and Hovey [2] generates the POD function by recording actual inspection response levels from cracks of varying size and plotting to determine the functional relationship between signal amplitude and crack size. The log/log (or log normal) function has been found to be representative of this relationship for many inspection processes. A sample plot of signal amplitude as a function of crack length is shown in Fig. 4. Using the pre-determined functional relationship between signal and flaw size, the data is transformed to a linear relationship and regression analysis is applied to determine a best fit line through the inspection data. Confidence bounds are calculated for the regression line and the inspection acceptance criteria is plotted as shown in Fig. 4. POD values as a function of crack length are determined by integrating the portion of the response distribution exceeding the acceptance threshold level for each crack length.

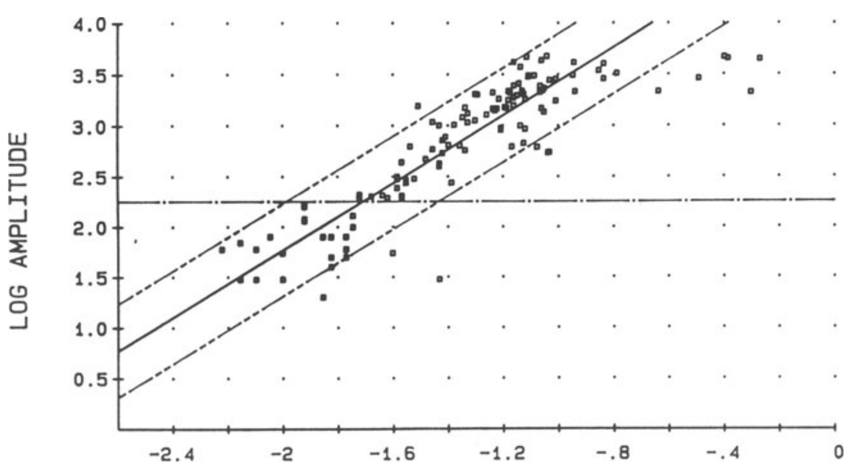

Fig. 4. Plot of signal amplitude as a function of crack length on a log/log scale for an eddy current hole inspection.

Assumptions that are inherent to the successful application of the â versus a method are (1) the response / size relationship can be modeled and made linear through data transformation; (2) the repetitive response distribution from a single crack and the response distribution from multiple cracks of equal size are normally distributed; and (3) the response variation distributions are equal for the range of flaw sizes sampled. A primary advantage of this analysis method is the ability to calculate POD curves at different acceptance threshold levels without the need to recollect data for each new threshold examined.

COMPARISON OF POD ANALYSIS METHODS

Data obtained from the assessment of the Intergrated Blade Inspection System (IBIS) fluorescent penetrant inspection module located at Kelly AFB, San Antonio, Texas [4] was used to plot POD curves using the three methods of analysis described above. The data was collected using turbine engine compressor blades containing 59 laboratory induced fatigue cracks of known sizes. The flawed specimens were inspected using production processing, calibration and data reduction procedures.

The resulting POD curves for a single inspection of the flawed compressor blade set using the binomial grouping, maximum likelihood and â versus a POD analysis methods respectively have been overlayed in Fig. 5 to illustrate the variation in results obtained with the three methods. As shown by these curves, the three analysis methods produced generally similar results.

A total of ten individual inspection sequences were completed on the fatigue flawed compressor blade set using the IBIS system. POD curves were plotted for each of these inspection sequences using the three analysis methods. The POD results for the ten inspections have been summarized in Table I as the crack lengths at which the mean and lower 95% confidence curves cross 90% probability of detection for the three methods of analysis. As shown by examination of the values in Table I, the largest variations in the POD curves resulting from the use of the different analysis methods were in the lower 95% confidence bounds. For the IBIS set of inspection capability data, the maximum likelihood method resulted in the most conservative lower 95% confidence estimate of POD and the binomial grouping method produced the least conservative lower 95% confidence level estimate. The mean or 50% confidence estimates of POD produced by the three methods exhibited much less variation than did the lower confidence curves. As a rule, the binomial grouping method produced the most conservative mean POD curves and the â versus a method the least conservative.

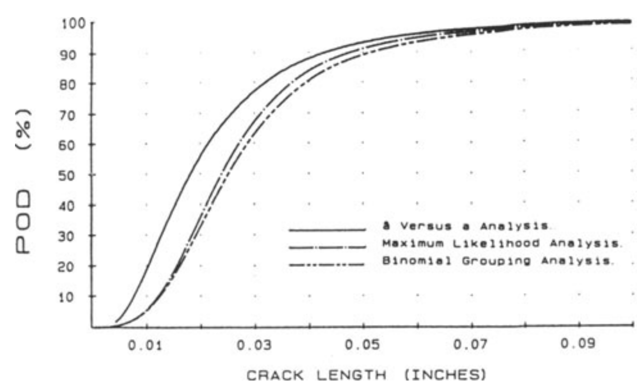

Fig. 5. Comparison of the mean POD curves for IBIS inspection sequence #7 using the three analysis methods.

TABLE I
Crack Lengths at 90% POD for Three Methods of Analysis using Mean and Lower 95% Confidence Lines

Insp.	â Versus a		Bin. Grouping		Max.Likelihood	
Number	90/50	90/95	90/50	90/95	90/50	90/95
1	.070 in.	.105	.067	.084	.075	.140
2	.062	.085	.067	.078	.064	.110
3	.058	.085	.070	.081	.063	.110
4	.058	.080	.063	.077	.063	.105
5	.053	.070	.057	.067	.052	.080
6	.045	.060	.051	.058	.048	.083
7	.064	.099	.078	.090	.080	.200
8	.053	.074	.060	.070	.048	.080
9	.053	.070	.065	.075	.059	.105
10	.065	.097	.060	.071	.052	.084

SPECIFICITY ANALYSIS

As mentioned previously, POD curves alone provide no information on the specificity or the ability of an inspection process to discriminate flaw signals from background noise. A high POD in conjunction with a high false call rate is not indicative of a practical production inspection process. Analysis of the data to determine the separation of flaw signals from background noise and the corresponding false call rate is necessary to fully understand the capability of an inspection process. Quantification of the relationship between POD and the probability of false alarms (POFA) is desirable to provide knowledge based options for management of an NDI system. A method which presents the relationship between POD and POFA as a function of the inspection acceptance threshold level is the specificity diagram. A sample specificity diagram for an eddy current hole inspection is shown in Fig. 6. This curve shows graphically the resulting POD and POFA over a range of possible acceptance level operating points. The availability of this curve allows the NDE engineer to select that acceptance criteria level that will result in the highest possible POD capability while still maintaining an acceptable false call level.

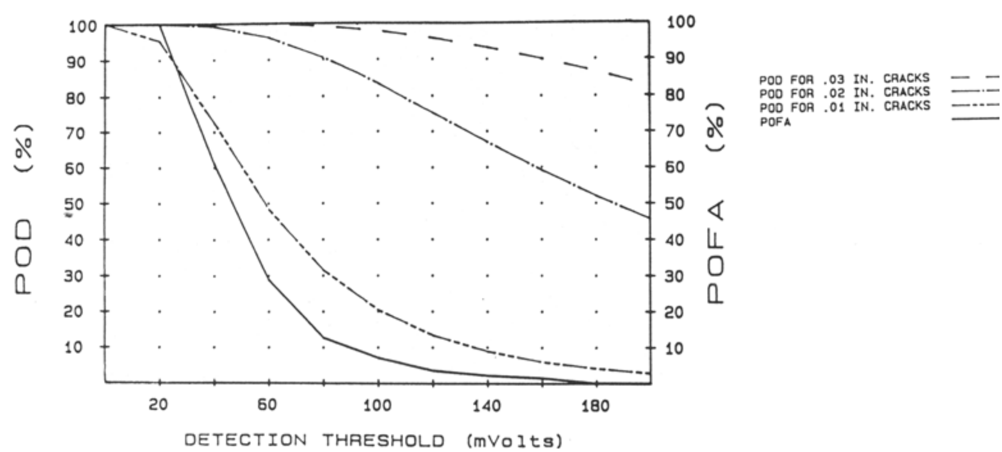

Fig. 6. Specificity diagram for an eddy current hole inspection procedure showing POD and POFA as a function of acceptance criteria.

SUMMARY AND CONCLUSIONS

Quantitatively knowing the detection capabilities of NDI processes is becoming increasingly important with the increased use of fatigue and fracture mechanics. This paper has presented some of the analysis methodology available for assessing and validating inspection techniques. Some of the topics discussed in detail include:

1. Inspection processes are a case of conditional probability and statistical hypothesis testing with four possible outcomes.

2. Output from an inspection is not absolute but is a distribution of signals that vary from inspection to inspection and flaw to flaw. Separation of the signal distribution from the noise distribution is required for a reliable inspection process.

3. The relationship between the signal and noise distributions and the acceptance criteria must be understood before a process can be operated in a production environment with confidence.

4. POD curves provide a convenient means of communicating the relationship between POD and flaw length. The $\hat{a}$ versus a, maximum likelihood and binomial grouping methods of of analysis commonly used to generate POD curves were described. A comparison of the results obtained with these methods was made using capability data generated using the IBIS penetrant inspection system. It was found that the methods produced generally similar results but as a rule, the $\hat{a}$ versus a method was the least conservative method and the binomial grouping method the most conservative.

5. Finally it was pointed out that a full specificity analysis of an inspection process is needed to correctly and quantitatively establish production operating parameters. The specificity diagram which presents POD and POFA as a function of acceptance threshold levels was presented. This diagram is an effective tool for aiding the NDE engineer in assessing the specificity of an inspection process and determining proper operating acceptance criteria levels.

ACKNOWLEDGEMENTS

The work described in this paper was sponsored in part by the Air Force San Antonio Air Logistics Center (SA-ALC) NDI Program Office by contract with Martin Marietta Astronautics Group.

REFERENCES

1. Rummel, Ward D., Brent K. Christner, Steve J. Mullen, and Donald L. Long, Characterization of Structural Assessment Testing, SA-ALC/MMEI/1/86, January, 1986.
2. Berens, A.P. and P.W. Hovey, Flaw Detection Reliablity Criteria, Volume 1 - Methods and Results, AFWAL-TR-84-4022, April, 1984.
3. Berkson, Joseph, M.D., Tables for the Maximum Likelihood Esitmate of the Logistic Function, Biometrics V.13, 1957.
4. Rummel, Ward D., Brent K. Christner, Steve J. Mullen, and Donald L. Long, Characterization of IBIS Fluorescent Penetrant Inspection Capabilities, SA-ALC/MMEI/3/86, November, 1986.